Time-Frequency Jigsaw Puzzle: Adaptive multiwindow and multilayered Gabor expansions

Florent Jaillet and Bruno Torrésani

***Abstract*— We describe a new adaptive family of multiwindow Gabor expansions, which adapt dynamically the windows to the signal's features in time-frequency space. The adaptation is based upon local time-frequency sparsity criteria, and also yields as by-product an expansion of the signal into layers corresponding to different windows. As an illustration, we show that using simply two different windows with different sizes leads to decompositions of audio signals into transient and tonal layers.**

***Index Terms*— Gabor expansion, multiwindow, time-frequency concentration, adaptivity, sparsity, entropy. EDICS: 2-TIFR - Non-stationary Signals, Time-Frequency and Frequency-Frequency Analysis**

I. Introduction

Time-frequency representations [2], [7] provide simple and efficient ways of representing signals. Among the various available time-frequency representations, the discrete Gabor transform (see [6] for a tutorial review of several aspects) has received considerable attention, from both theoretical and applied points of view. It essentially provides expansions of signals as linear combinations of *time-frequency atoms*, with fixed time and frequency "concentration" properties.

A classical problem in time-frequency analysis, and particularly in Gabor analysis, is that of adapting the representation to the analyzed signal. Using sparsity criteria, we showed that global adaptations do not make sense, except for very simple one-component signals [9], [10]. In general, such an optimization is possible only in limited regions of the time-frequency plane (where signal components are isolated [9]).

The recently proposed multiwindow Gabor expansion (see e.g. [16], [5]) offer more flexibility, by using several windows in the same scheme. The inherent increase of redundancy of the representation makes the analysis at the same time more complex and richer, as it allows more flexibility in the expansion of the signal (i.e. it provides many more different ways to reconstruct the signal from coefficients). However, it may be relaxed by a suitable (signal dependent) reduction of the so-created dictionary of waveforms.

We describe here an adaptive way of reducing this dictionary, by solving a "Time-Frequency Jigsaw Puzzle" problem: pave the time-frequency plane with "super-tiles", within which the "locally optimal" Gabor expansion is retained. This is performed by exploiting local (in time and frequency) sparsity criteria, namely entropy-like measures. within a given region (super-tile) of the time-frequency plane, the time-frequency atoms which yield maximally sparse signal representation are retained. As a result, we obtain adaptive Gabor expansions, together with a corresponding multilayered signal decomposition, in which different layers are provided by different windows.

The approach we develop here is in many respects similar to the recently proposed schemes for nonlinear approximation in dictionaries of waveforms (including among others matching pursuit and orthogonal matching pursuit, basis pursuit,...). However, the main difference here is that one of the goals of our approach is to produce decompositions of signals into significantly different layers, which leads us to use dictionaries made out of unions of frames with significantly different properties[1].

Both authors are with LATP, CMI, 39 rue Joliot-Curie, 13453 Marseille cedex 13, France; F. Jaillet is also with GENESIS S.A., Batiment Beltram, Domaine du petit Arbois, BP 69, F-13545 Aix en Provence Cedex 4, France; email: jaillet@genesis.fr; Bruno.Torresani@cmi.univ-mrs.fr. Work supported in part by the European Union's Human Potential Programme, under contract HPRN-CT-2002-00285 (HASSIP)

II. Multiwindow Gabor expansions

A. Multiwindow and reduced multiwindow expansions

Let us start by briefly sketching the main features of Gabor expansions that will be of interest for the present discussion. Let $g \in L^2(\mathbb{R})$, $g \neq 0$, and $b, \nu \in \mathbb{R}^+$, and consider the set of functions

$$g_{mn} = e^{2i\pi n\nu t} g(t - mb) \ , \quad m, n \in \mathbb{Z} \ , \tag{1}$$

naturally associated with the lattice $\mathcal{L} = b\mathbb{Z} \times \nu\mathbb{Z}$ in the time-frequency plane. It may be shown (see for example [3]) that for $b\nu$ small enough, the family of time-frequency atoms $\{g_{mn}, m, n \in \mathbb{Z}\}$ constitute a *frame* in $L^2(\mathbb{R})$: there exist constants $0 < A \leq B < \infty$ such that for all $f \in L^2(\mathbb{R})$,

$$A\|f\|^2 \leq \sum_{m,n\in\mathbb{Z}} |\langle f, g_{mn}\rangle|^2 \leq B\|f\|^2 \ .$$

It then follows from the general frame theory that there exist inversions formulas for the *coefficient map* $f \in L^2(\mathbb{R}) \rightarrow \{\langle f, g_{mn}\rangle, m, n \in \mathbb{Z}\} \in \ell^2(\mathbb{Z}^2)$, of the form

$$f = \sum_{m,n\in\mathbb{Z}} \langle f, g_{mn}\rangle \, h_{mn} \ , \tag{2}$$

for some "dual" function h. In general, such an inverse is far from unique, which raises the question of finding the "optimal" h for a given g (or conversely the optimal g for a given h.) The canonical dual window, denoted hereafter by $\tilde{g}$, is obtained by considering the inverse of the *frame operator* $\mathcal{R}$, defined by

$$\mathcal{R}x = \sum_{m,n\in\mathbb{Z}} \langle x, g_{mn}\rangle g_{mn} \ , \tag{3}$$

[1] Such dictionaries may be used as well in matching pursuits and orthogonal matching pursuits, but the application to multilayered signal decompositions does not seem to have been considered in that context.

in which case one has

$$\tilde{g} = \mathcal{R}^{-1} g \ . \tag{4}$$

It is a classical issue that for a given g, all corresponding "Gaborlets" g_{mn} have the same time-frequency concentration properties (since they are time and frequency shifted copies of each other, up to a phase factor), which is not necessarily convenient for practical purposes. Multiwindow Gabor expansions [16] were proposed as a simple alternative, providing a way to adapt time-frequency resolution to the signal. We stick here to a simple version.

Given R Gabor frames $\mathcal{F}_r = \{g^r_{mn}, m, n \in \mathbb{Z}\}$, $r = 0, \dots R-1$ the system $\mathcal{F} = \mathcal{F}_0 \cup \mathcal{F}_1 \cup \dots \cup \mathcal{F}_{R-1}$ is obviously still a frame of $L^2(\mathbb{R})$. Any signal x may thus be reconstructed from the coefficients

$$\alpha^r_{m,n} = \langle x, g^r_{mn} \rangle \ ,$$

such an inverse transform being far from unique (for example, any weighted average of the individual expansions would do the job, these being certainly not the most interesting solutions). We shall rather be interested in (signal dependent) *reduced* expansions [5], that involve limited number of (dual) time-frequency atoms. We shall be concerned here with adaptive redundancy reduction algorithms.

Remark 1: simple multiwindow Gabor systems. A simple way to generate such multiwindow Gabor systems is to start from a reference Gabor frame $\mathcal{F} = \{g_{mn}, m, n \in \mathbb{Z}\}$, with corresponding time-frequency lattice $\mathcal{L} = b\mathbb{Z} \times \nu\mathbb{Z}$, generate various dilated copies of the window g: $g^r(t) = \alpha^{-r/2} g(\alpha^{-r} t)$ and adapt the time-frequency lattice $\mathcal{L}_r = \alpha^r b\mathbb{Z} \times \alpha^{-r}\nu\mathbb{Z}$ accordingly. While we shall stick to the simple case $R = 2$, we develop the scheme below in the general case. Notice also that many other variations are possible.

Remark 2: multilayered signal decomposition. One of the goals of this work is to provide decompositions of signals into *layers*, that are characterized by a few (large) coefficients in one of the frames of the multi Gabor system. Before going into the details of the proposed algorithm and the description of the selection criteria, let us first describe a simple approach to achieve this goal, in a simple situation. Suppose that, starting from two Gabor frames for simplicity, some criterion has selected two corresponding sets of time-frequency atoms $\{g^0_\lambda, \lambda \in \Lambda\}$ and $\{g^1_\delta, \delta \in \Delta\}$, where Λ and Δ are two subsets of the index sets associated with g^0 and g^1. Assume that these two sets are frames of the subspace of $L^2(\mathbb{R})$ they span; then the corresponding frame operators, denoted by $\mathcal{R}_{0,\Lambda}$ and $\mathcal{R}_{1,\Delta}$ and defined as follows: for $x \in L^2(\mathbb{R})$,

$$\mathcal{R}_{0,\Lambda} x = \sum_{\lambda \in \Lambda} \langle x, g^0_\lambda \rangle \, g^0_\lambda \ , \qquad \mathcal{R}_{1,\Delta} x = \sum_{\delta \in \Delta} \langle x, g^1_\delta \rangle \, g^1_\delta \tag{5}$$

are invertible on their ranges. Introducing the corresponding dual time-frequency atoms[2]

$$\gamma^0_\lambda = \mathcal{R}^{-1}_{0,\Lambda} g^0_\lambda \ , \lambda \in \Lambda \ , \qquad \gamma^1_\delta = \mathcal{R}^{-1}_{1,\Delta} g^1_\delta \ , \delta \in \Delta \ , \tag{6}$$

we can build the orthogonal projection operators on the subspaces spanned by the two frames: for all $x \in L^2(\mathbb{R})$,

$$\mathcal{P}_{0,\Lambda} x = \sum_{\lambda \in \Lambda} \langle x, g^0_\lambda \rangle \, \gamma^0_\lambda \ , \quad \mathcal{P}_{1,\Delta} x = \sum_{\delta \in \Delta} \langle x, g^1_\delta \rangle \, \gamma^1_\delta \ . \tag{7}$$

Then, for all $x \in L^2(\mathbb{R})$, we can write

$$x = \mathcal{P}_{0,\Lambda} x + \mathcal{P}_{1,\Delta} x + r \ , \tag{8}$$

where r is some residual signal, equal (up to a minus sign) to the orthogonal projection of x onto the intersection $Span(\{g^0_\lambda, \lambda \in \Lambda\}) \cap Span(\{g^1_\delta, \delta \in \Delta\})$. The latter equation indeed provides a decomposition of the signal x into two layers, plus some residual to be processed further: it may be dispatched "democratically" into the two layers, or expanded similarly into the two Gabor frames, starting from new suitably adapted index sets Λ and Δ.

This approach is *not* the approach we follow below, mainly for the following reason: it involves the computation of *restricted* dual frames $\{\gamma^0_\lambda \lambda \in \Lambda\}$ and $\{\gamma^1_\lambda \delta \in \Delta\}$, which is not necessarily easy (nor numerically stable, see [5]) for arbitrary index sets Λ and Δ. In addition, the explicit dependence of the dual time-frequency atoms on the index sets makes it necessary to recompute them at eact step if an iterative scheme is used to process the residual, which yields high computational burden.

B. Adaptive redundancy reduction: TFJP1

Let us now state the reduction problem in terms of sparsity requirements. We first introduce the sparsity criterion we shall be using. For the sake of simplicity, we limit ourselves to a finite version of a Rényi entropy, but let us point out that other criteria (other Rényi entropies, or the Shannon entropy) may be used as well (see [9], [14] for a discussion of this point). We first choose $\alpha \in (0, 1)$. Given a finite vector $u \in \mathbb{C}^N$, with norm $\|u\| = \sqrt{\sum_0^{N-1} |u_n|^2}$, we set

$$R_\alpha(u) = \frac{1}{1-\alpha} \log_2 \left[\sum_{n=0}^{N-1} \left(\frac{|u_n|^2}{\|u\|^2} \right)^\alpha \right] \ . \tag{9}$$

It is very easy to see that R_α is indeed a measure of sparsity, as it is maximal for constant vectors (up to a phase factor), and minimal for "Kronecker-like" vectors $u_n = C\delta_{n,n_0}$.

Let us start from a multiwindow Gabor expansion as before, with windows $g^1, \dots g^R$, time-frequency lattices $\mathcal{L}^0, \dots \mathcal{L}^{R-1}$, and corresponding canonical dual windows[3] $\tilde{g}^1, \dots \tilde{g}^R$, we fix $A, B \in \mathbb{R}^+$ large enough, and consider a reference tiling of the time-frequency plane into rectangular "super-tiles"

$$\mathbb{R}^2 = \bigcup_s \square_{(s)} \tag{10}$$

where the super-tiles are defined by

$$\square_{(s)} = \square_{k,n} = [kA, (k+1)A) \times [nB, (n+1)B) \ . \tag{11}$$

We are interested into different pavings of these by time-frequency atoms. For each time-frequency lattice $\mathcal{L}^r$, let

[2] In this paper, we shall reserve the "tilde" notation for canonical dual time-frequency atoms constructed using the "full" frame operator $\mathcal{R} = \mathcal{R}_{\mathbb{Z}^2}$, and use another notation form frames defined in terms of subsets of the index sets. Notice also that we have removed the explicit dependence on the index set in the notation, so as to avoid too heavy notations.

[3] Notice that we use here the canonical dual windows, associated with the full time-frequency lattices $\mathbb{Z}^2$, thus avoiding the problems mentioned in the previous section.

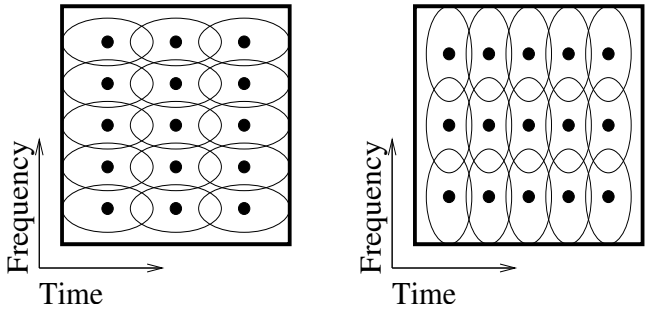

Fig. 1. Two different lattices within a super-tile; the rectangular region is the super-tile, the ellipses represent the "numerical" support of the time-frequency atoms (say, the domain within which their spectrogram exceeds some fixed threshold), and the dots represent their center, i.e. the time-frequency sampling points.

$\mathcal{L}^{r,s} = \mathcal{L}^r \cap \square_{(s)}$ denote the subset of $\mathcal{L}^r$ included in the super-tile $\square_{(s)}$. An illustration of a super-tile and corresponding pavings and sub-lattices may be found in Figure 1.

To any $x \in L^2(\mathbb{R})$, any window g^r and any super-tile s, associate the set of coefficients $\langle x, g^r_{mn} \rangle$ corresponding to the paving of super-tile s with atoms of type r: $\alpha(x,r,s) = \{\langle x, g^r_{mn}\rangle,\ m,n \in \mathcal{L}^{r,s}\}$. Then, for each super-tile s, the optimal window $g^{r(x,s)}$ and the corresponding entropy are defined by

$$\left\{ \begin{array}{ll} r(x,s) & = \arg\min_{r=0,\dots R-1} H(\alpha(x,r,s)) \ , \\ H(x,s) & = H(\alpha(x,r(x,s),s)) \ . \end{array} \right. \tag{12}$$

Given $x \in L^2(\mathbb{R})$, let the first approximation be defined as

$$x^{(1)} = \sum_s \sum_{\lambda \in \mathcal{L}^{r(x,s),s}} \langle x, g^{r(x,s)}_\lambda \rangle \ \tilde{g}^{r(x,s)}_\lambda \tag{13}$$

and the corresponding residual

$$\mathcal{R}^1(x) = x - x^{(1)} \ . \tag{14}$$

The procedure may then be iterated:

$$x^{(k)} = \sum_s \sum_{\lambda \in \mathcal{L}(x,k,s)} \langle \mathcal{R}^{k-1}(x), g^{r(x,k,s)}_\lambda \rangle \ \tilde{g}^{r(x,k,s)}_\lambda \ , \tag{15}$$

$$\mathcal{R}^k(x) = \mathcal{R}^{k-1}(x) - x^{(k)} \ . \tag{16}$$

where $r(x,k,s) = r(\mathcal{R}^{k-1}(x), s)$ corresponds to the window selected at step k within the supertile s, and $\mathcal{L}(x,k,s) = \mathcal{L}^{r(x,k,s),s}$ denotes the corresponding time-frequency sampling points.

At step K, we then obtain a telescopic expansion of the signal into K approximation levels and a residual

$$x = \sum_{k=1}^K x^{(k)} + \mathcal{R}^K(x) \ , \tag{17}$$

and the iteration stops when the residual is small enough.

Remark 3: convergence issues. The theoretical study of the convergence of such a scheme turns out to be more complex than that of matching pursuit type algorithms (see e.g. [8]), the decision criterion being more complex itself. Nevertheless, numerical illustrations shown below seem to indicate exponential convergence for suitable choices of the parameters (windows, sampling grids,...).

III. Multilayered time-frequency decomposition

We now change our point of view, and propose a different way of analyzing the result of the method. For the sake of simplicity, we limit the discussion to the case $R = 2$, i.e. the case of multigabor systems with two windows only. The extension to more than 2 windows is straightforward. The iterative algorithm described above also yields directly multilayered signal decompositions, as described below. In numerical experiments, we shall limit ourselves to the particular case of two identical windows, at two different scales: a wide version and a narrow version. This choice is motivated by the desire of decomposing audionumeric signals into "tonal" and "transient" layers. In this spirit, the tonal layer of a signal is defined as the "component" which admits a sparse expansion with respect to a Gabor frame with high frequency resolution (i.e. with a wide window), and the transient layer as the "component" which admits a sparse expansion with respect to a Gabor frame with high time resolution (i.e. a narrow window).

A. Multilayered decomposition in the context of TFJP1

Given an expansion of the type (17), each approximation level is itself expressed as a linear combination of Gabor atoms with different window functions: setting, for $\rho = 0, 1$, $S(x,k,\rho) = \{s : r(x,k,s) = \rho\}$, we may write

$$\mathcal{R}^k(x) = \sum_{\rho=0}^1 \sum_{s \in S(x,k,\rho)} \sum_{\lambda \in \mathcal{L}(x,k,s)} \langle \mathcal{R}^{k-1}(x), g^\rho_\lambda \rangle \ \tilde{g}^\rho_\lambda \ , \tag{18}$$

which rewrites as

$$x = \ell_0(x) + \ell_1(x) + \mathcal{R}^K(x) \ , \tag{19}$$

where the remainder $\mathcal{R}^K(x)$ is as before, and

$$\ell_\rho(x) = \sum_{k=1}^K \sum_{s \in S(x,k,\rho)} \sum_{\lambda \in \mathcal{L}(x,k,s)} \langle \mathcal{R}^{k-1}(x), g^\rho_\lambda \rangle \ \tilde{g}^\rho_\lambda \ . \tag{20}$$

The layer $\ell_0(x)$ (resp. $\ell_1(x)$) basically represents the "component" of the signal x which is "well represented" (i.e. sparsely represented) by the Gabor frame $\mathcal{F}^0$ (resp. $\mathcal{F}^1$). Actually, if the two windows have sufficiently different characteristics (in particular, time-frequency localization properties), the different layers do indeed represent very significantly different components of the signal. Numerical applications on audio signals are shown below.

In the particular case of two windows, the flow chart of the algorithm is given in Figure 2. Even though the description of the algorithm may appear a bit "tricky", we wish to point out that its structure is in fact quite simple.

B. A simple variant: TFJP2

The approach described above treats all layers equally, in the sense that at each iteration, the construction of the time-frequency puzzle is followed directly by the simultaneous estimation of corresponding contributions to all layers. However it turns out that in such a scheme, the estimate of the tonal layer

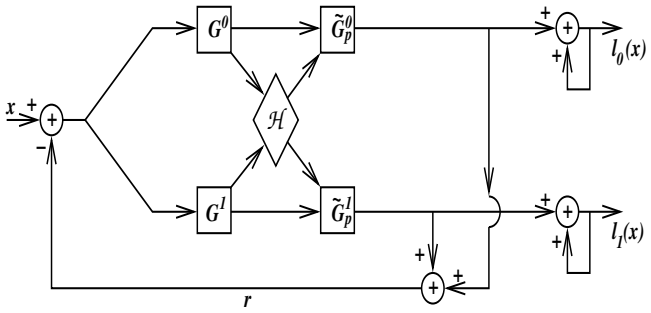

Fig. 2. Diagram of the proposed algorithm in the case of two windows. G^0 and G^1 represent the analysis maps for the two Gabor frames, and $\tilde{G}_p^0$ and $\tilde{G}_p^1$ represent the partial synthesis maps from the selected time-frequency atoms. $\mathcal{H}$ represents the calculation of the entropy and the corresponding decision.

may be affected by the presence of the transient one, and vice-versa. To avoid such shortcomings, it is possible to modify slightly the algorithm, and only estimate a single layer at each iteration. More precisely, assume for the sake of simplicity that two windows are given. Given a signal x, the first step is still given in the same way as (12)

$$\left\{ \begin{array}{ll} r(x,s) & = \arg\min_{r=0,1} H(\alpha(x,r,s)) \ , \\ H(x,s) & = H(\alpha(x,r(x,s),s)) \ , \end{array} \right. \tag{21}$$

but the first estimate only takes into account the first window

$$x^{(1;0)} = \sum_{s:r(x,s)=0} \sum_{\lambda \in \mathcal{L}^{0,s}} \langle x, g_\lambda^0 \rangle \ \tilde{g}_\lambda^0 \tag{22}$$

which also defined the corresponding residual

$$\mathcal{R}^{\frac{1}{2}}(x) = x - x^{(1;0)} \ . \tag{23}$$

The second window is then used to estimate the contribution to the second layer: Equation (21) is used again, and the second estimate reads

$$x^{(1;1)} = \sum_{s:r(\mathcal{R}^{\frac{1}{2}},s)=1} \sum_{\lambda \in \mathcal{L}^{1,s}} \langle \mathcal{R}^{\frac{1}{2}}(x), g_\lambda^1 \rangle \ \tilde{g}_\lambda^1 \ . \tag{24}$$

The residual is then

$$\mathcal{R}^1(x) = \mathcal{R}^{\frac{1}{2}}(x) - x^{(1;1)} = x - x^{(1,0)} - x^{(1;1)} \ . \tag{25}$$

Again, the procedure may be iterated, taking the residuals $\mathcal{R}^k$ as inputs: replacing x with $\mathcal{R}^k$ in Equations (21) to (24) yields similarly

$$x^{(k;0)} = \sum_{s:r(\mathcal{R}^{k-1}(x),s)=0} \sum_{\lambda \in \mathcal{L}^{0,s}} \langle \mathcal{R}^{k-1}(x), g_\lambda^0 \rangle \ \tilde{g}_\lambda^0 \tag{26}$$

$$x^{(k;1)} = \sum_{s:r(\mathcal{R}^{k-\frac{1}{2}}(x),s)=1} \sum_{\lambda \in \mathcal{L}^{1,s}} \langle \mathcal{R}^{k-\frac{1}{2}}(x), g_\lambda^1 \rangle \ \tilde{g}_\lambda^1 \ , \tag{27}$$

and residuals

$$\mathcal{R}^{k-\frac{1}{2}}(x) = \mathcal{R}^{k-1}(x) - x^{(k,0)} \ , \tag{28}$$

$$\mathcal{R}^{k}(x) = \mathcal{R}^{k-\frac{1}{2}}(x) - x^{(k;1)} \ . \tag{29}$$

As a result, one obtains a telescopic series

$$x = \sum_{k=1}^{K} \left(x^{(k,0)} + x^{(k,1)} \right) + \mathcal{R}^K(x) \tag{30}$$

as well as two layers

$$\ell_0(x) = \sum_{k=1}^{K} x^{(k,0)} \tag{31}$$

$$\ell_1(x) = \sum_{k=1}^{K} x^{(k,1)} \ , \tag{32}$$

with the same interpretation as before.

As we shall see in Section IV, this variant has the advantage of better avoiding boundary effects between adjacent super-tiles in which different windows are chosen. It also yields slightly better convergence.

C. Introducing significance test for sparsity: TFJP1b

The main idea of the above algorithms is to choose, within each super-tile s, the window such that the resulting entropy is minimal. However, the minimal entropy for a given super-tile may happen to be quite large, meaning that for that particular super-tile, even the "best" window was unable to yield a sufficiently sparse description. In such situations, it does not necessarily make sense to include the contribution of the considered super-tile in one of the layers, an alternative being to keep it inside the residual. We describe below this new approach (TFJP1b) in the framework of the TFJP1 algorithm (the modifications needed to adapt it to the TFJP2 algorithm are straightforward).

For a given super-tile s, and corresponding values of entropies $H(\alpha(x,r,s))$, one has to decide whether or not those values are significant (i.e. correspond to actual significant signal component.) To avoid possible non-significant values, we decide that the optimal window defined in (12) is accepted only when the corresponding entropy is below some threshold value. Given such a threshold $\tau \in \mathbb{R}^+$, we simply replace (13) and (14) with

$$x_\tau^{(1)} = \sum_{s:H(x,s)\leq\tau} \sum_{\lambda \in \mathcal{L}^{r(x,s),s}} \langle x, g_\lambda^{r(x,s)} \rangle \ \tilde{g}_\lambda^{r(x,s)} \tag{33}$$

$$\mathcal{R}_\tau^1(x) = x - x^{(1)} \ , \tag{34}$$

and similarly for the rest of the algorithm. The multilayered Gabor expansion may also be adapted accordingly, within the scheme depicted in Figure 2. This now produces a decomposition of the signal into three layers: the two previous ones, and a residual.

Remark 4: choice of the threshold. In some specific applications, the threshold τ may of course be chosen by the user. In a more general context, it may be desirable to choose the value(s) of the threshold on statistical grounds, which is however difficult, as it would require characterizing the distribution of Shannon's entropies computed from restrictions of Gabor transforms to super-tiles.

In our numerical experiments to be discussed below, we used the following procedure. The distribution of the entropies was estimated (numerically) from Gabor coefficients of a white noise reference signal. τ was then adjusted to a given significance level (for example, 5%). In other words, at each step of the iterative algorithm, super tiles were rejected (and kept in the residual signal) when the corresponding value of

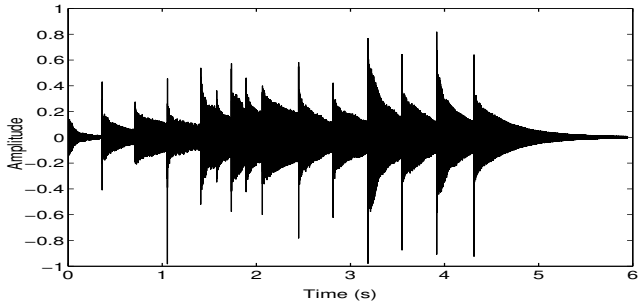

Fig. 3. Glockenspiel test signal

entropy was too likely to have been produced by a Gaussian white noise (considered the worst case signal, as far as spartity is concerned). In such situations, the residual has no reason to converge to zero, and may even contain interesting signal which simply cannot be sparsely represented by the considered systems of time-frequency atoms.

IV. Numerical results

We present below a number of numerical experiments that illustrate the behavior of the proposed algorithms from various points of view: the speed of convergence of the iterative algorithm, the speed of convergence of the approximations obtained by retaining the largest coefficients, and an application of the multilayered decomposition to speech processing.

In all the numerical experiments, we have limited ourselves to one decision criterion: the Renyi entropy R_α, with $\alpha = 0.5$ (see 9). Similar results are obtained with different criteria.

A. Convergence issues

In order to study the convergence of the proposed algorithms, we tested them on three sample signals, which represent different levels of difficulty for the algorithm:

- "SinDir": a sum of a sine wave and a "Dirac" pulse
- "Noise": realization of zero-mean, white noise (with uniform pdf).
- "Glock": real audio signal, namely a *Glockenspiel* signal, displayed in Fig. (3).

These sample signals were tested in the following configurations: two Gaussian windows of different size (and bandwidth) were used:

- Experiment A, window sizes of 128 and $128 \times 5 = 640$ samples.
- Experiment B, window sizes of 128 and $128 \times 33 = 4224$ samples.

In both experiments, the three supertile configurations were tested:

- Configuration 11: the time-frequency super-tiles correspond to 1 time sampling point with the wide window, and 1 frequency sampling point with the narrow window.
- Configuration 33: the time-frequency super-tiles correspond to 3 time sampling points with the wide window, and 3 frequency sampling points with the narrow window (as in Fig. 1).

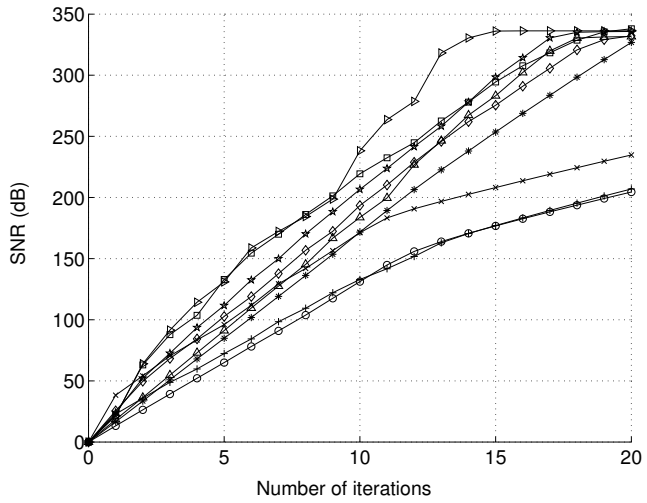

Fig. 4. Convergence of the TFJP1 algorithm in a few situations in experiment A (window sizes of 128 and 640 samples respectively), for the three sample signals (Noise, SinDir and Glock) in the three considered configurations (11, 33 and 55); Signal to Noise Ratio as a function of the number of iterations. $\circ$: Noise11; $\times$: SinDir11; $+$: Glock11; $*$: Noise33; $\square$: SinDir33; $\diamond$: Glock33; Δ: Noise55; $\triangleright$: SinDir55; $\star$: Glock55.

- Configuration 55: the time-frequency super-tiles correspond to 5 time sampling points with the wide window, and 5 frequency sampling points with the narrow window.

We first describe numerical results obtained with TFJP1. As may be seen from Fig. 4 and Fig. 5, the convergence is very satisfactory in all situations, a signal to noise ratio larger than 100dB being obtained in less than 20 iterations. The 11 configuration appears to yield slower convergence, presumably because it is the one that produces the most important boundary effects between supertiles (configurations 33 and 55 correspond to larger super-tiles, and therefore produce less boundaries). Also, faster convergence is observed in experiments B, i.e. when the two windows are more significantly different (except in configuration 11). As may be expected also, the "Noise" signal yields slower convergence, because one does not expect it to have any sparse expansion in the considered dictionary; similarly, faster convergence is observed for the SinDir signal, which is specially taylored for this algorithm. Notice that in Experiment B (larger "wide window"), the numerical precision of the calculations (about 320 dB) is reached in less than 15 iterations (except for the 11 configuration, which seem less favourable, as stressed above).

Similar results for two versions of TFJP2 are displayed in Figures 6 and 7. These two versions differ in the choice of the window g^0 which is used first. In Figure 6, g^0 is the wide window, and in Figure 7, g^0 is the narrow window.

The main result is that the very first iterations provide better accuracy in this situation. This is specially neat for the SinDir and the Glock signal, less spectacular for the noise signal. This effect is more pronounced in Figure 6 (wide window used first). This is a consequence of the nature of the signals (SinDir and Glock), in which the harmonic components have larger energy than the transient ones.

We also compared the performances of TFJP1 and TFJP2 in terms of convergence of the approximation, again using a

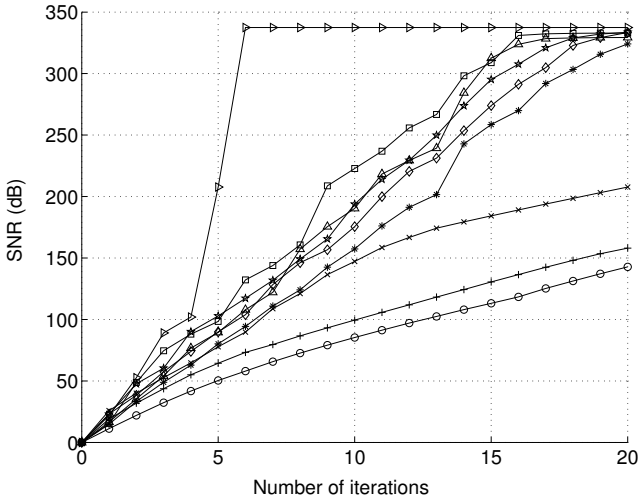

Fig. 5. Convergence of the TFJP1 algorithm in a few situations in experiment B. Same legend as Fig 4.

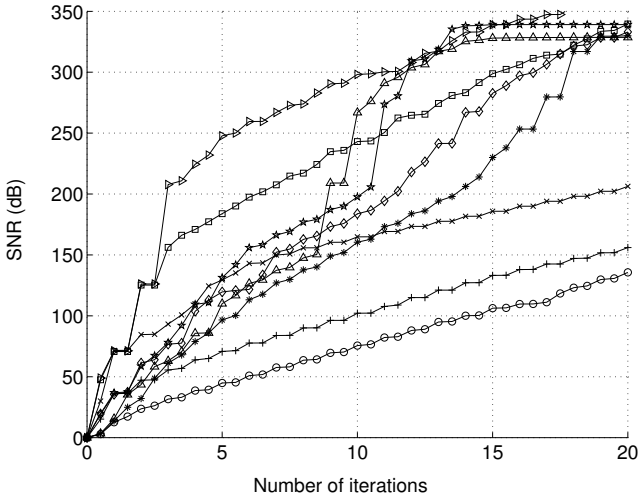

Fig. 6. Convergence of the TFJP2 algorithm in a few situations in experiment B, with g^0 the wide window. Same legend as Fig 4.

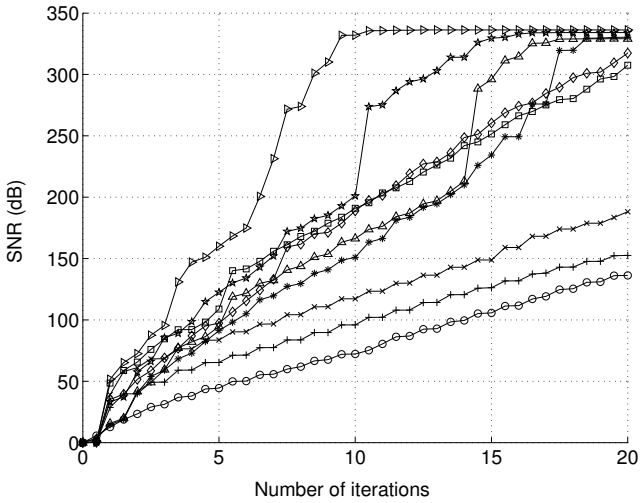

Fig. 7. Convergence of the TFJP2 algorithm in a few situations in experiment B, with g^0 the narrow window. Same legend as Fig 4.

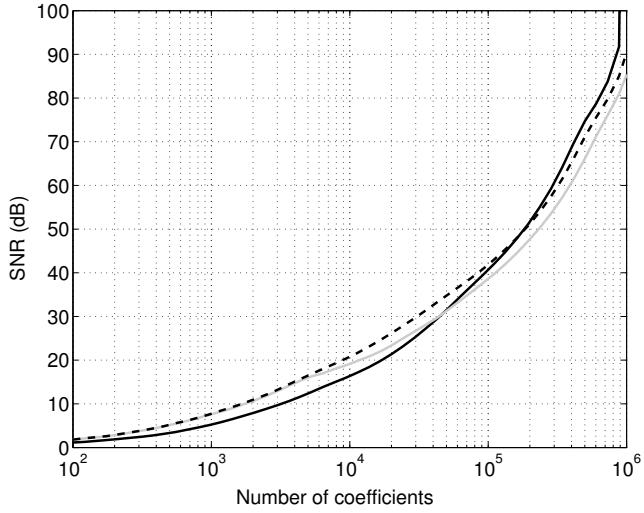

Fig. 8. Comparison of TFJP1, TFJP2 and a "standard" Gabor representation in terms of convergence of N-term approximations: SNR as a function of the number of retained coefficients. Full black line: standard Gabor expansion; dashed line: TFJP1; full gray line: TFJP2.

wide and a narrow Gaussian window. In addition to TFJP1, two occurences of TFJP2 are considered: in the first one, g_0 is the wide window and g_1 the narrow one, and vice versa in the second one. A first element of comparison is obtained by looking at the time-frequency representations of the so-obtained tonal and transient layers (not shown here). As may be expected, the tonal components are better resolved in the first case, while the transients are better resolved in the other case. TFJP1 provides intermediate results.

To compare these versions, we also considered N-term approximations, i.e. truncated sums

$$x_N = \sum_{n=0}^{N-1} \alpha_n g_n^{i_n}$$

where the considered atoms are chosen in such a way that the corresponding coefficients are the N largest ones in magnitude. We considered the Glockenspiel signal in Fig. 3, and tested TFJP1 and TFJP2. We display in Fig. 8 the corresponding values of the SNR as a function of the number of coefficients used for the reconstruction. For the sake of comparison, we also display the same quantity for a "standard" Gabor representation, using a window of intermediate size.

As may be seen, for the same number of retained coefficients, the SNR is larger by a few dBs for TFJP2. Both TFJP1 and TFJP2 outperform the classical Gabor approach, as long as the number of retained coefficients is not too large. For larger number of coefficients, the result is different, presumably because the TFJP methods involve twice more atoms than the classical Gabor one.

B. Transient-Tonal decomposition

A natural field of application of the algorithm we just described is provided by audiophonic signals, and the simultaneous estimation of transient and tonal layers (also considered in [4], [13].) As explained above, the starting point is to

define the tonal layer of a signal as the "component" which admits a sparse expansion with respect to a Gabor frame with high frequency resolution (i.e. with a wide window), and the transient layer as the "component" which admits a sparse expansion with respect to a Gabor frame with high time resolution (i.e. a narrow window). As before, we considered two windows: a wide window g^0 and a narrow window g^1 (in fact two copies of a Gaussian window at different scales), and considered the multilayered expansions of (19) and (20) for various examples of audio signals.

We display in Figure 9 the time-frequency representations of the two layers (transient and tonal) obtained by TFJP2 on the Glockenspiel signal, and the time-frequency representation of the complete signal. As may be seen on the three images, the separation of the two components is extremely neat. The corresponding waveforms are available on the web site attached to this paper (see the conclusion).

To illustrate the decomposition into three layers (tonal, transient and residual), we display in Figure 10 the results of the decomposition obtained using TFJP1b on a speech signal: the word `/test/`. Remarkably enough, the algorithm was able to separate the different letters of the signal: the `t` are captured by the transient layer, the `e` by the tonal layer, and the `s` remain in the residual. However, such results turn out to be quite sensitive to shifts of the super tiles. Therefore, a systematic exploration of such approaches for speech signal processing will require extra tuning effort, which we plan to study in the future.

V. Conclusions, perspectives

We have presented in this paper a new approach for automatic selection of adapted Gabor signal representations, starting for several "standard" Gabor expansions with different window functions. This approach is quite general, and may be adapted in various ways. We have also presented some of these variations.

As a by-product, this approach also yields "multilayered" representations for the signal under study, a layer being defined as the "component" of the signal that is well represented by a given type of Gabor functions.

Even though we have focused here on a few illustrations on general audio and speech signals and transient/tonal separation, we believe that such approaches possess a much wider application range. To quote only a few of these, applications to blind source separation or automatic speech segmentations are examples of applications which we plan to address in the near future.

Additional material, including additional figures, and sound files, may be found on a companion web site:
`http://www.cmi.univ-mrs.fr/~torresan/papers/TFJP`

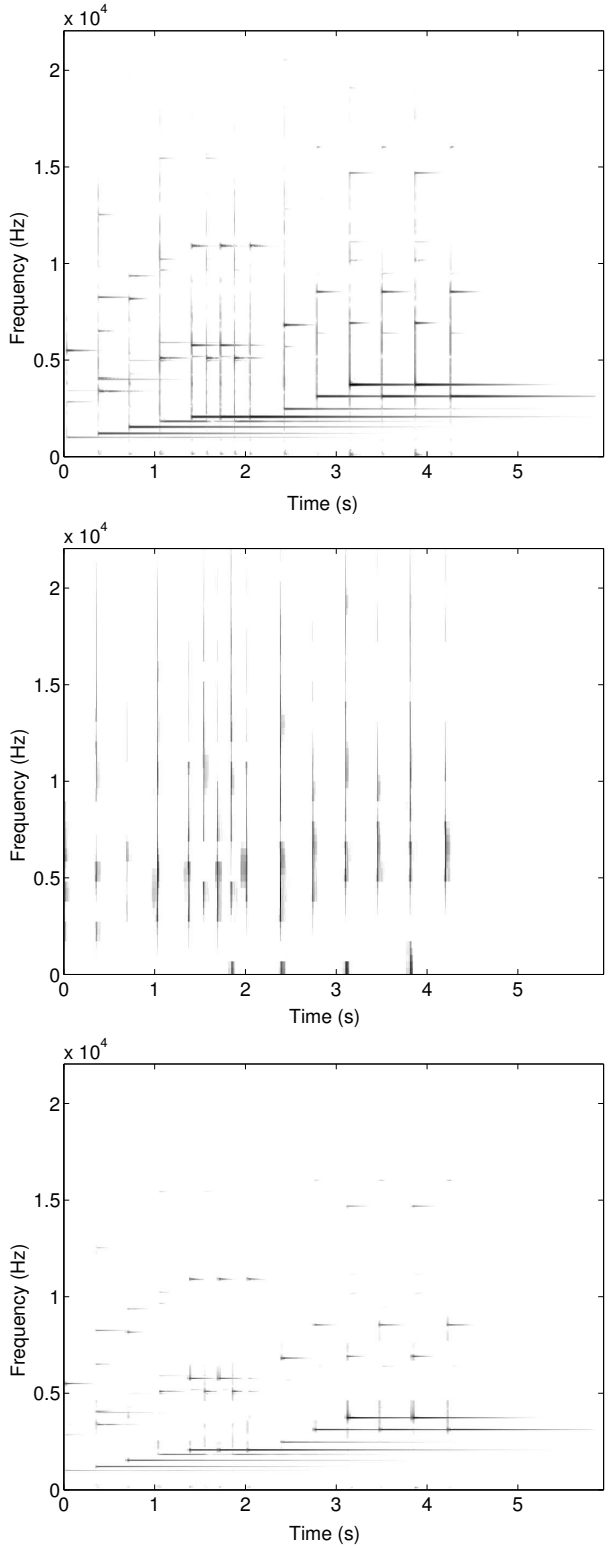

Fig. 9. Multilayered decomposition of the "Glockenspiel" signal, obtained using TFJP2. From top to bottom: time-frequency representations (spectrograms) of the original signal (with a "medium size" window), the estimated transient layer (with the narrow window) and the estimated tonal layer (with the wide window).

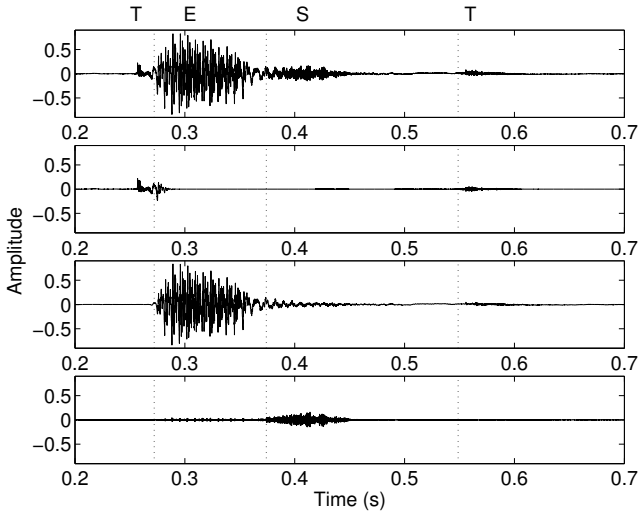

Fig. 10. Multilayered decomposition of a short piece of speech signal: `/test/`, obtained using TFJP1b. From top to bottom: waveforms of the original signal, the transient layer, the tonal layer and the residual signal.

REFERENCES

[1] J. Berger, R. Coifman, and M. Goldberg. Removing noise from music using local trigonometric bases and wavelet packets. *J. Audio Eng. Soc.*, 42(10):808–818, 1994.

[2] R. Carmona, W.L. Hwang, and B. Torrésani. *Practical Time-Frequency Analysis: continuous wavelet and Gabor transforms, with an implementation in S*, volume 9 of *Wavelet Analysis and its Applications*. Academic Press, San Diego, 1998.

[3] I. Daubechies. *Ten Lectures on Wavelets*, volume 61 of *CBMS-NFS Regional Series in Applied Mathematics*, 1992.

[4] L. Daudet and B. Torrésani. Hybrid representations for audiophonic signal encoding. *Signal Processing*, 82(11):1595–1617, 2002. Special issue on Image and Video Coding Beyond Standards.

[5] M. Dörfler, *Gabor Analysis for a Class of Signals called Music*, PhD thesis, NuHAG, Vienna, 2003.

[6] H.G. Feichtinger and T. Strohmer, *Gabor analysis and algorithms, theory and applications*, Birkhäuser Boston Inc., 1998.

[7] P. Flandrin, *Time-Frequency/Time-Scale Analysis*, volume 10 of *Wavelet Analysis and its Applications*. Academic Press, San Diego, 1999.

[8] R. Gribonval. *Approximations non-linéaires pour l'analyse des signaux sonores*. PhD thesis, Université de Paris IX Dauphine, 1999.

[9] F. Jaillet and B. Torrésani. *Remarques sur l'adaptativité des représentations temps-fréquence*. Proceedings of the symposium *GRETSI'03*, Paris, France, 2003.

[10] F. Jaillet and B. Torrésani. *Adaptive time-frequency representation for sound analysis and processing*. Proceedings of the symposium *Congrès Français d'Acoustique*, Strasbourg, France, 2004.

[11] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41:3397–3415, 1993.

[12] S. Molla and B. Torrésani. Hidden markov trees of wavelet coefficients for transient detection in audiophonic signals. In A. Benassi, editor, *Proceedings of the conference Self-Similarity and Applications, Clermont-Ferrand (May 2002)*, 2003. to appear.

[13] S. Molla and B. Torrésani. An Hybrid Audio Scheme using Hidden Markov Models of Waveforms *Applied and Computational Harmonic Analysis*, to appear.

[14] A. Trgo and M.V. Wickerhauser. A relation between Shannon–Weaver entropy and "theoretical dimension" for classes of smooth functions. Preprint, Washington University, Saint Louis, Missouri, 1995.

[15] M. V. Wickerhauser. *Adapted Wavelet Analysis from Theory to Software*. AK Peters, Boston, MA, USA, 1994.

[16] M. Zibulski and Y. Zeevi, Analysis of multiwindow gabor-type schemes by frame methods, *Applied and Computational Harmonic Analysis*, **4**, Vol. 2 (1997), pp. 188-212.